

A Sample Information Retrieval Index

Document A : Now is the time.

Document B : For all good men.

Document C: To come to the aid of their country.

	Document A	Document B	Document C
Now	1		
Is	1		
The	1		1
Time	1		
For		1	
All		1	
Good		1	
Men		1	
To			2
Come			1
Aid			1
of			1
their			1
country			1

Figure 1

CLAUSE
 NP (SUBJ)
 The [article]
 Department of Justice [noun (phrasal, gov_entity)]
 VP (ACTIVE_VOICE)
 sued [verb, past tense]
 NP (DOBJ,)
 Microsoft [noun, (company)]
 PP
 for [preposition]
 NP
 antitrust [adjective]
 violations [noun]
 PP
 in
 NP
 federal [adjective]
 court [noun]

Figure 2a

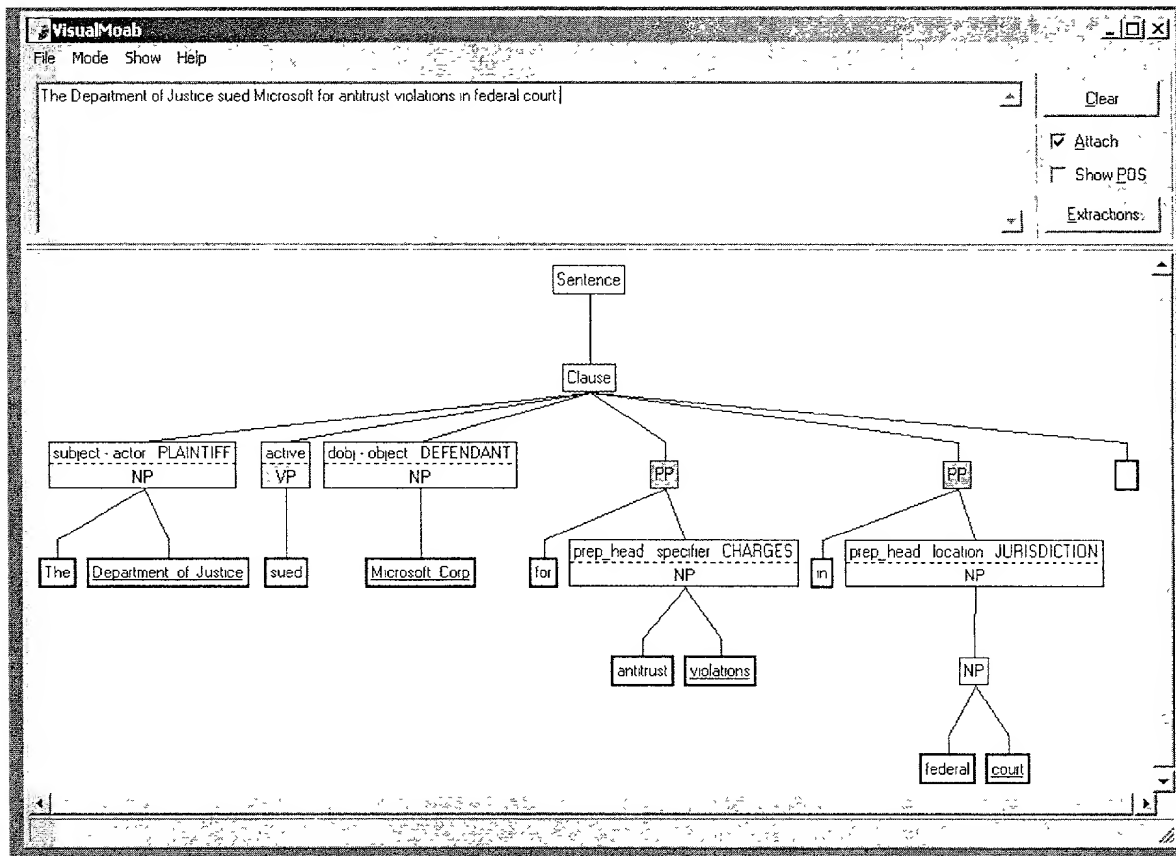


Figure 2b
A Graphical View of A Sentence Parse and Thematic Role Assignment

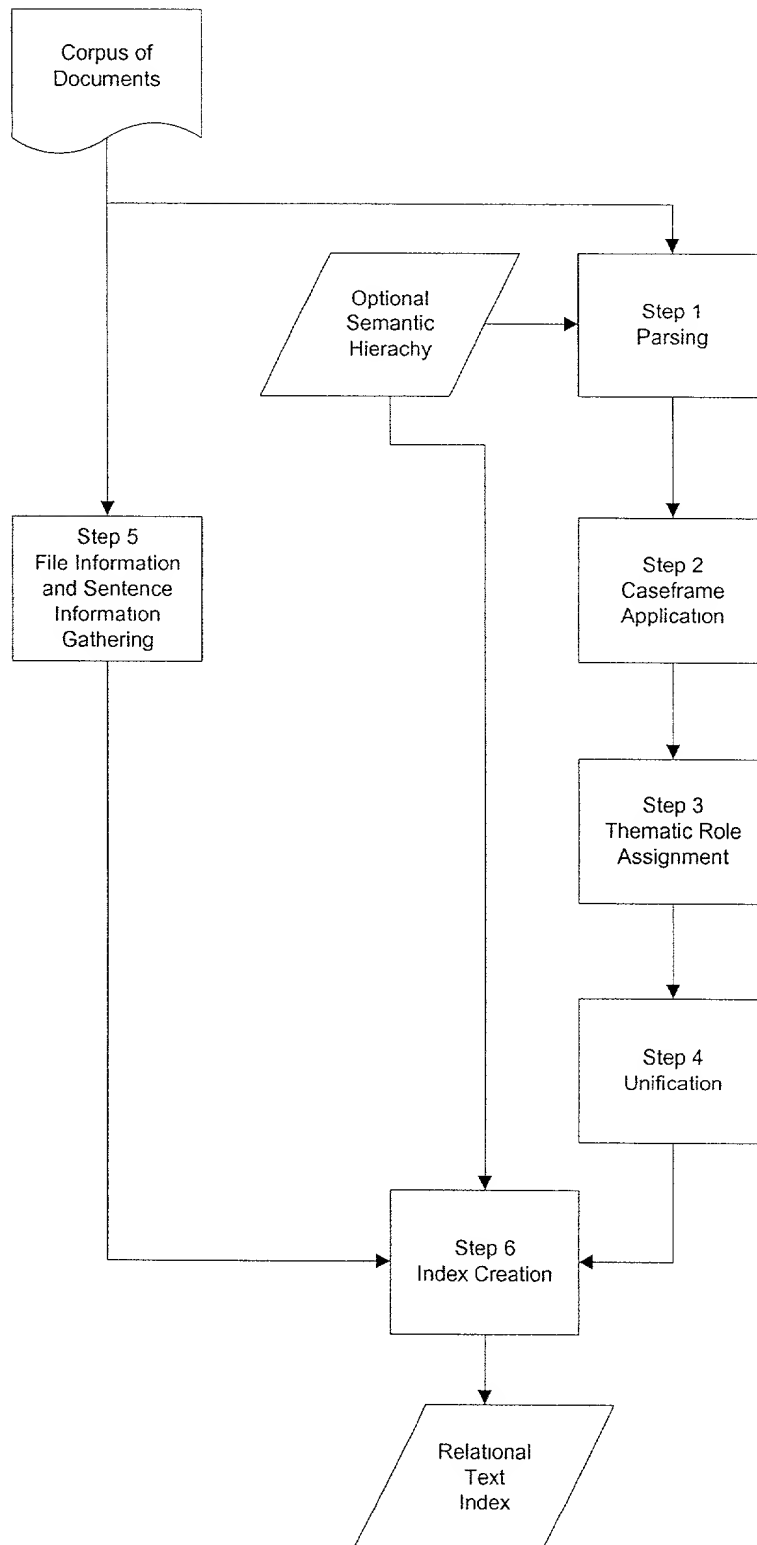


Figure 3
Index Creation Algorithm, High Level

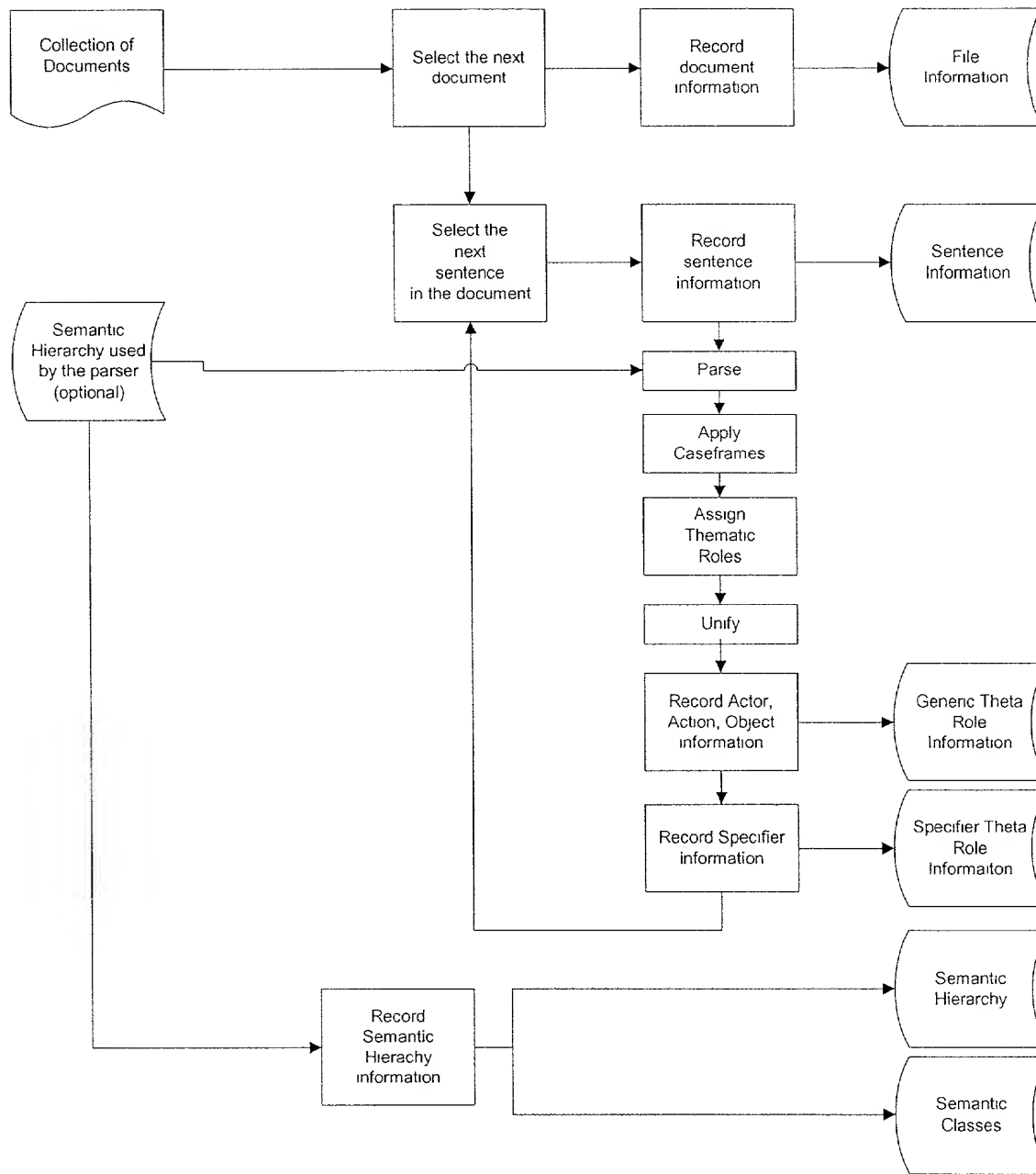


Figure 4
Index Creation Algorithm, Low Level

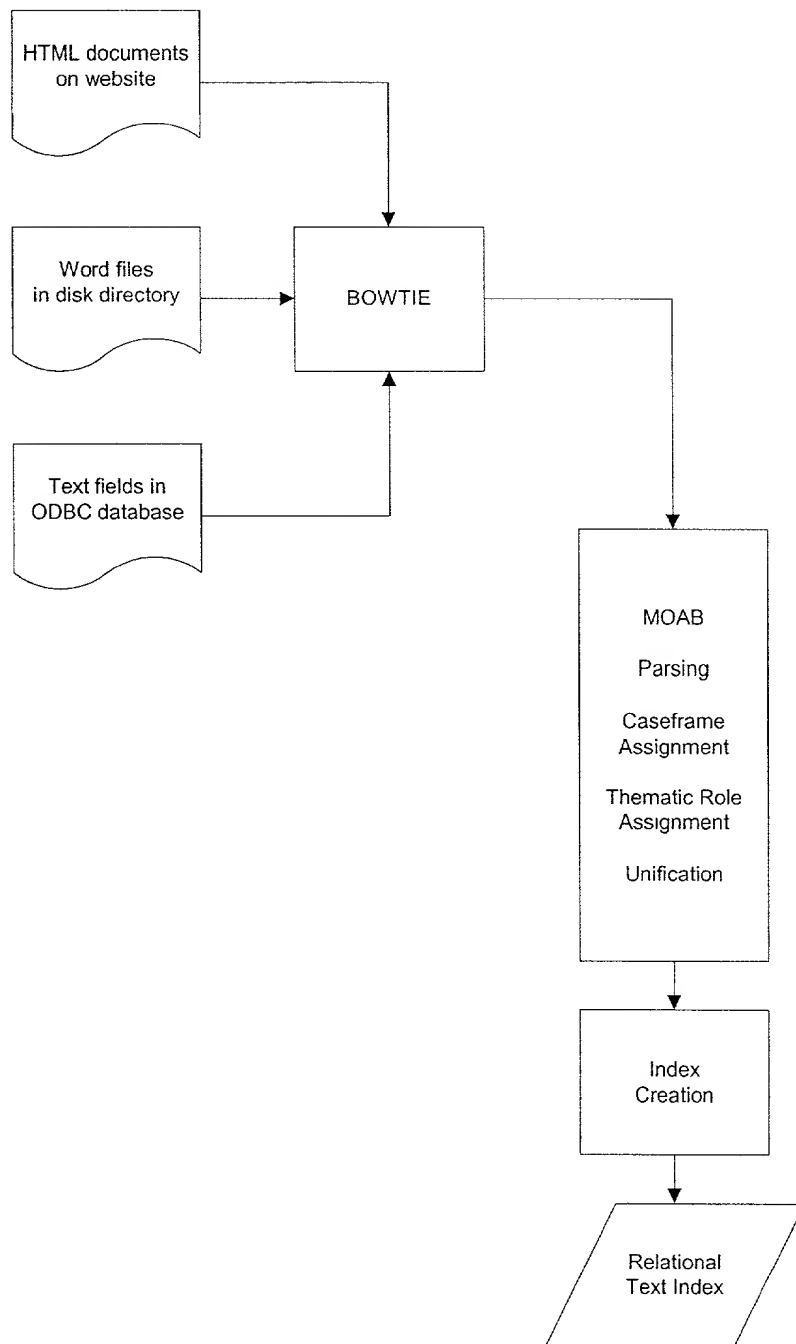


Figure 5
Overall Flow

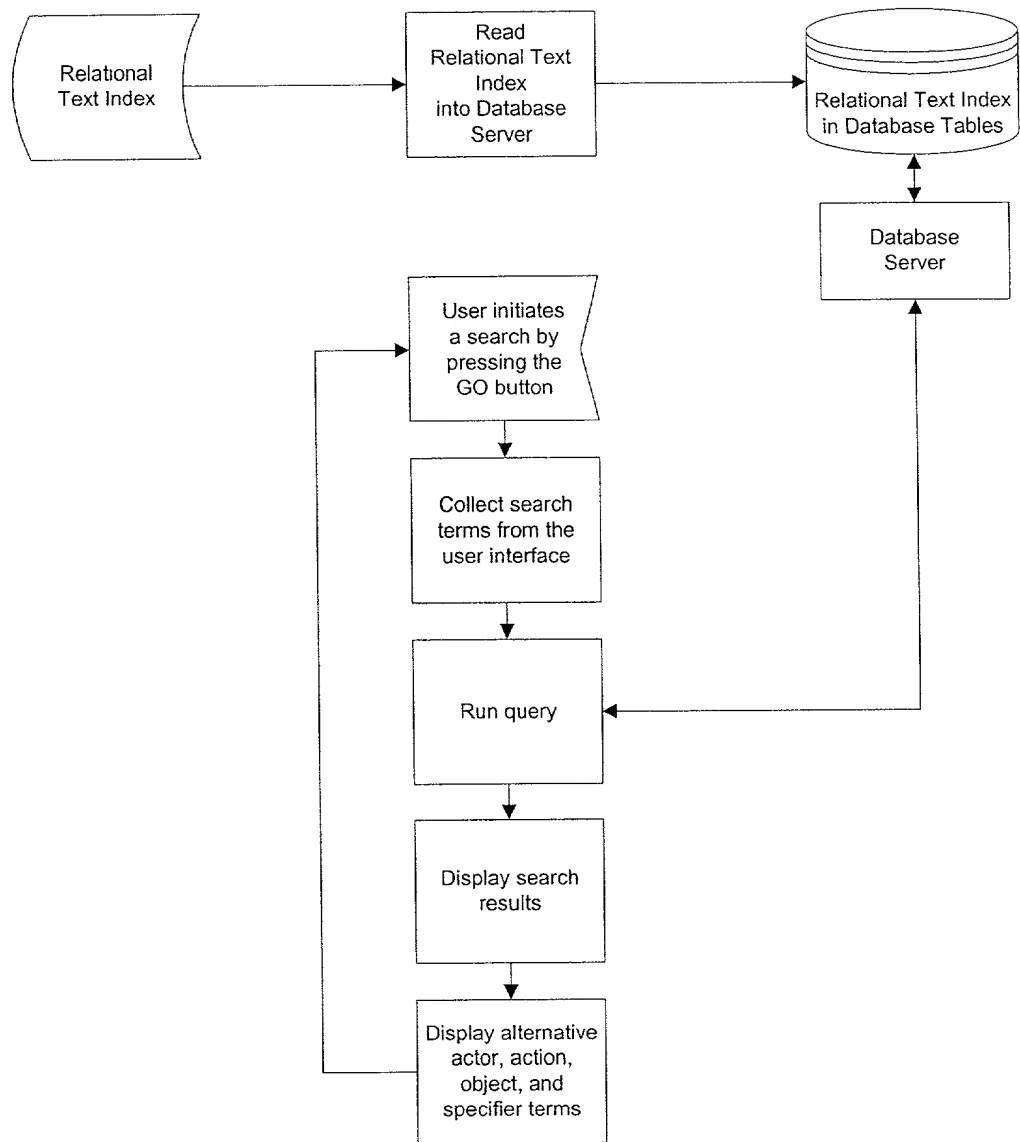


Figure 6
Flowchart – PowerDrill

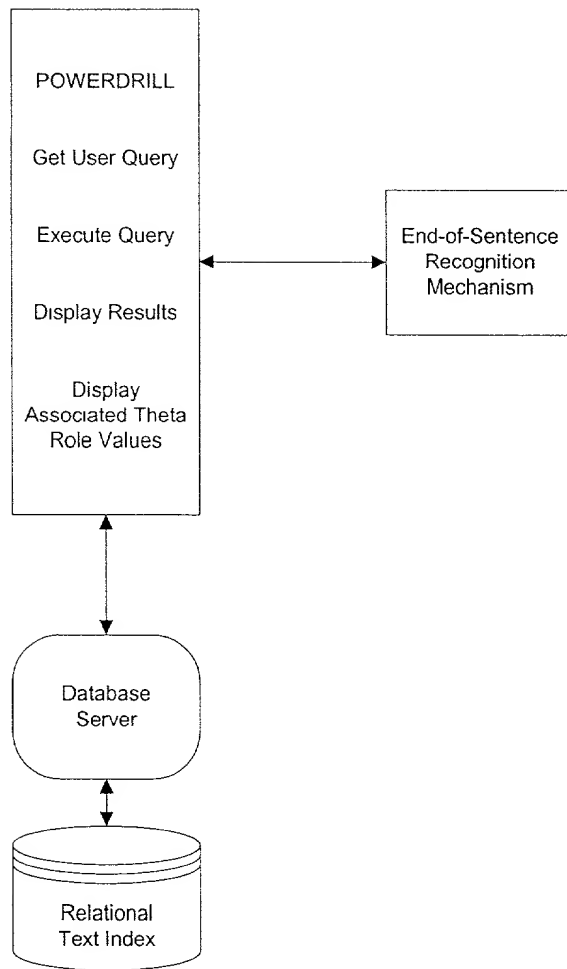


Figure 7
Overall Search Processing

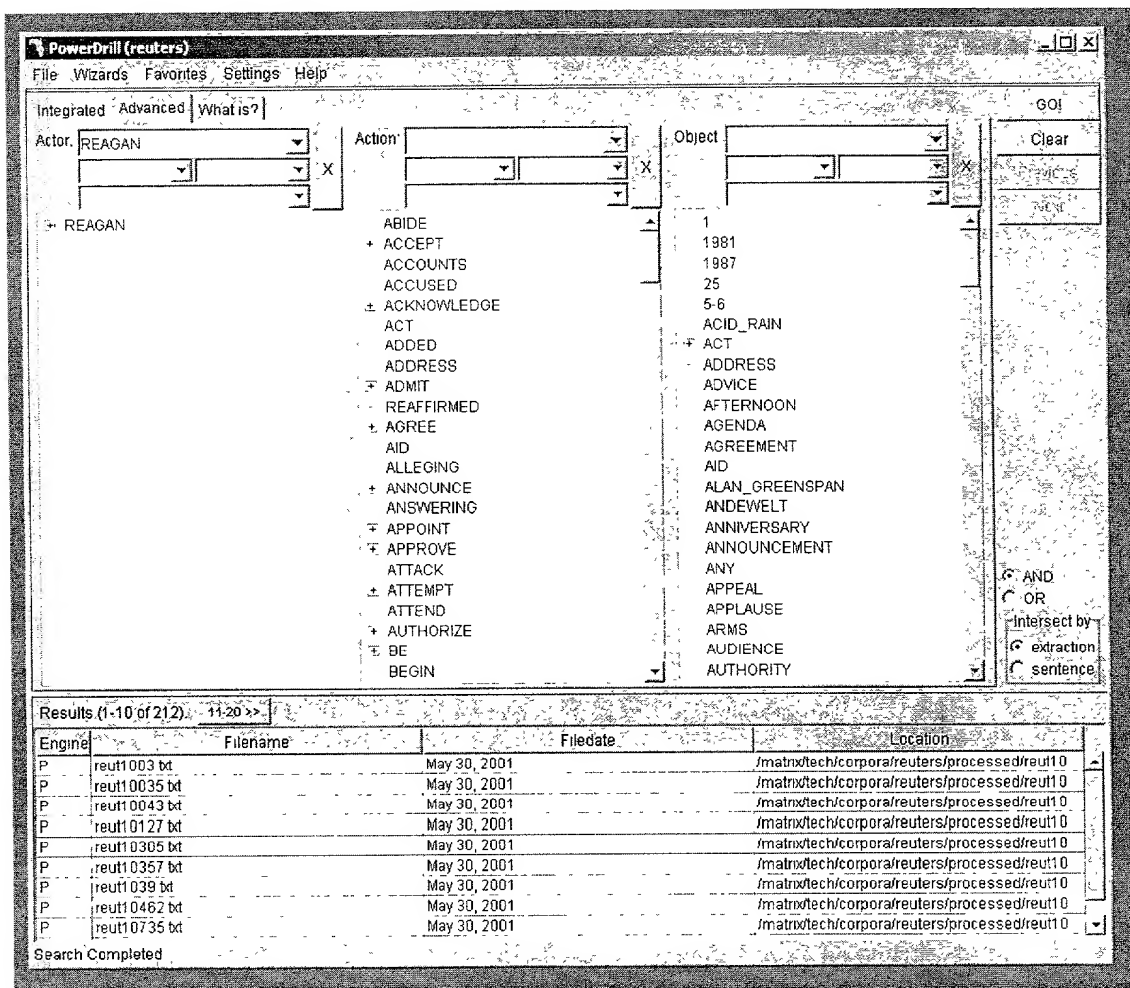


FIGURE 8

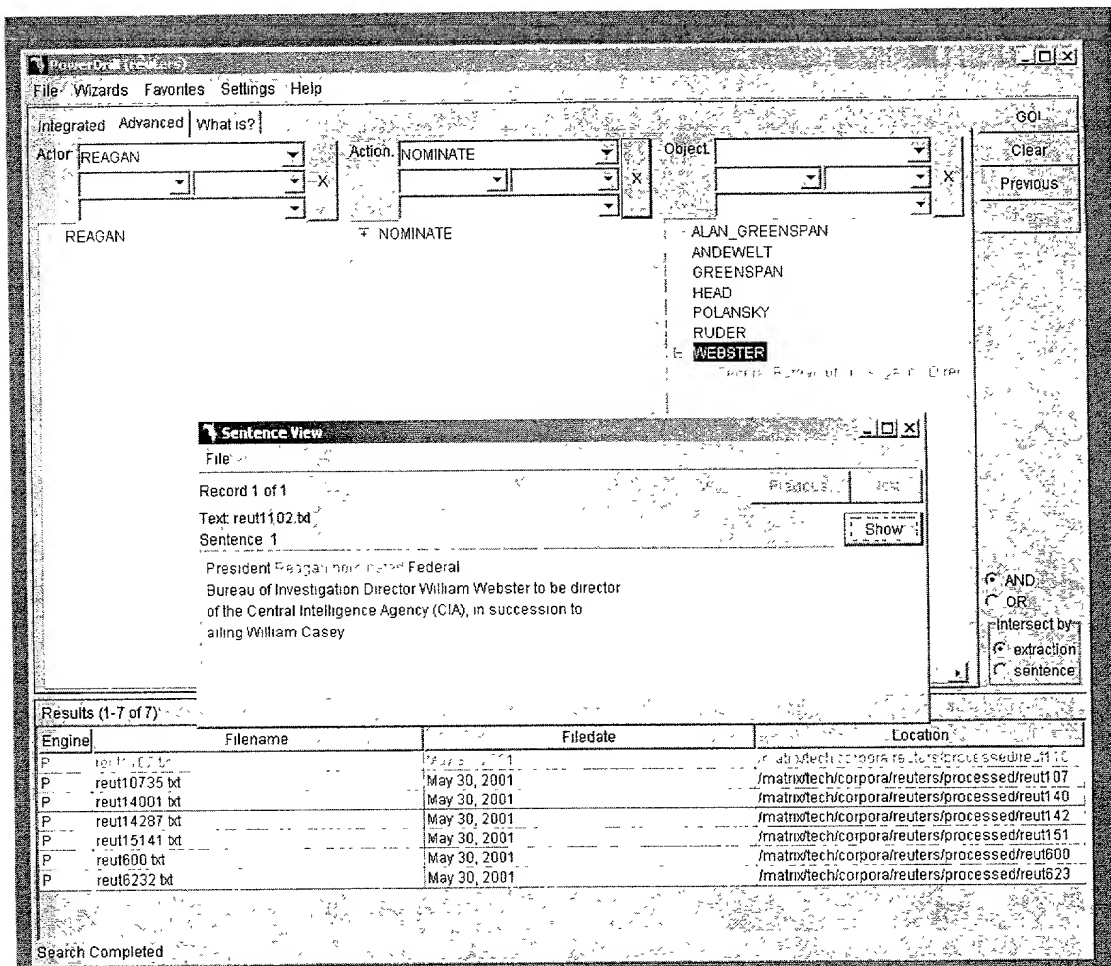


FIGURE 9

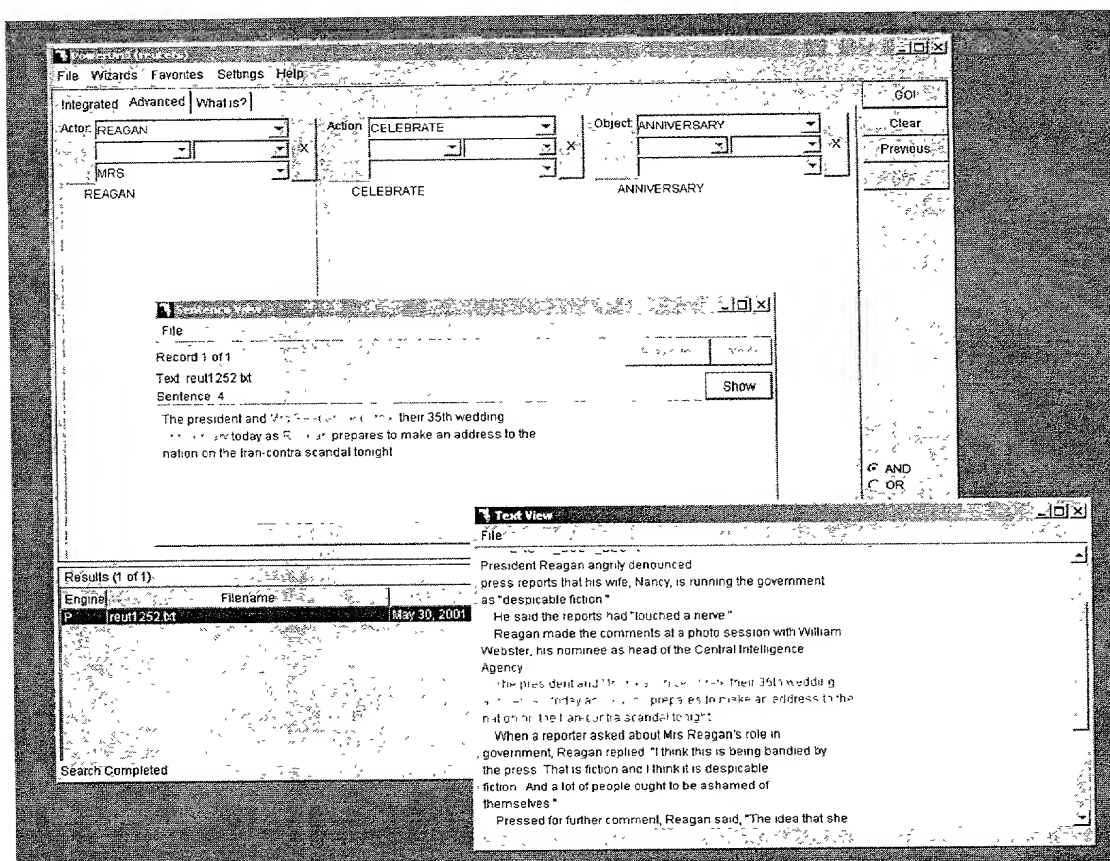


Figure 10

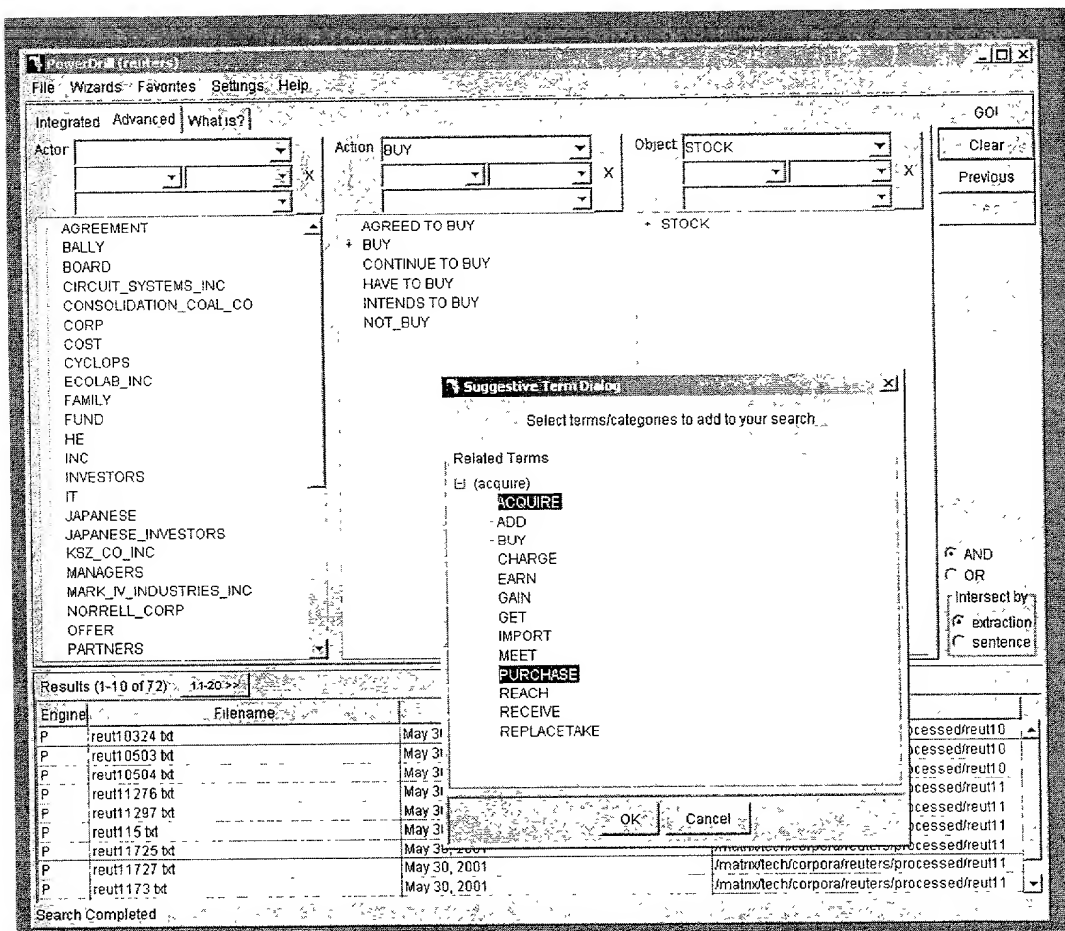


Figure 11

PowerDrill (reuters)

File Wizards Favorites Settings Help

Integrated: Advanced What is?

Actor

INVESTOR

X

Action

BUY ACQUIRE PURCHASE

X

Object

X

GO!

Clear

Previous

INVESTOR

ACQUIRE

AGREE TO BUY

APPEAR TO BUY

BUY

ENTITLE

FLOCKED TO BUY

PURCHASE

AMOUNT

BENEFITTED

BOND ISSUES

BULLION

CERTIFICATES

COMMON_SHARE

CONVERTIBLE

DEBT

DOLLARS

FLORIDA

ISSUES

MARK

MEADOWS

METAL

PERCENTAGE

SECURITIES

SHARE

SILVER

STOCKS

THEM

UNITED_STATES

VESTOR

WARRANT

AND

OR

Intersect by

extraction

sentence

Results (1-10 of 29) 11:20 >>

Engine	Filename	Filedate	Location
P	reut12504 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12542 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12543 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12544 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12545 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12546 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12547 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12548 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12549 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11
P	reut12550 bt	May 30, 2001	/matrxtech/corpora/reuters/processed/reut11

Search Completed

Figure 12

PowerDrill (reuters)

File Wizards Favorites Settings Help

Integrated Advanced Whatis?

Actor: [Bank] X Action: VIOLATE X Object: LAW X

BANK
BLY
COMPANY
DEFENDANTS
FIRST_FEDERAL
HE
IT
PERSONS
RESOLUTION
SALE
SHARES
THRIFT
TRANS_WORLD_AIRLINES_INC

- LAW
- LAW
- LAWS

AND
OR
Intersect by:
extraction
sentence

Results (1-10 of 55) < 11:20 >

Engine	Filename	Filedate	Location
P&K	reut7604.txt	May 30, 2001	/mainwtech/corpora/reuters/processed/reut760
For Federal Circuit court, Bank of America National Association, et al., v. ... violated ... law ...			
The lower court and First Federal violated ... law ...			
P&K	reut4533.txt	May 30, 2001	/mainwtech/corpora/reuters/processed/reut453
... violated ... law ...			
The ... violated ... law ...			
... violated ... law ...			

Search Completed

Figure 13